DØ Note 5066-CONF

Z + jet production in the DØ experiment: A comparison between data and the PYTHIA and SHERPA Monte Carlos

The DØ Collaboration
URL: http://www-d0.fnal.gov
(Dated: March 21, 2006)

A first comparison between DØ data and predictions from the event generators SHERPA and PYTHIA has been performed for the Z/γ^* + jet production with $Z/\gamma^* \rightarrow e^+e^-$. Data corresponding to an integrated luminosity of about 950 pb^{-1} have been collected between October 2002 and November 2005 by the DØ experiment at the upgraded Fermilab Tevatron collider. The prediction from the SHERPA Monte Carlo, combining parton shower and matrix element calculations, has been found to give an accurate description of jet multiplicities. In addition, the p_T spectra of the Z boson and of the jets as well as angular correlations between jets are reasonably well described.

Preliminary result for Winter 2006 Conferences

I. INTRODUCTION

The study of the associated production of a vector boson with jets represents an important test of QCD at hadron colliders. In addition, W/Z + jet production constitutes an important background in the search for many new physics processes, e.g. in the Higgs boson search in the associated WH and ZH production. A precise knowledge of the total and differential production cross sections is essential for a reliable background estimate in these searches.

The most accurate approach to describe a multi-particle final state would be to include all particles in a full matrix element computation including all real and virtual diagrams. However, the number of final state particles which can be described in this way is limited both by the rapid growth of the number of diagrams to be evaluated, and by the complexity of each single diagram. In general, two simplifications can be made:

- one can generate a 2→ 2 core process, and describe all other final state particles as being emitted from the core particles in a probabilistic manner using a parton shower approach. This simplification is valid in the limit where all emissions are soft/collinear.
- one can include real emission diagrams in the matrix element computation and ignore virtual corrections. To avoid soft/collinear divergencies one has to introduce cut-offs which limit the topology of the final state.

During the last few years several partially overlapping approaches for combining the two methods have been proposed, one of them being the CKKW [1, 2] algorithm. The idea is to generate 2→ N processes by including only tree-level diagrams, i.e. diagrams without loops, in the matrix element computations. Cuts in the phase space of the final state particles are introduced to avoid the soft/collinear divergences. Thereafter one uses a parton shower on the N final state particles to populate the rest of the phase space. In this way the hard and well-separated jets will be described by the matrix element, keeping interference terms, whereas the soft/collinear jets are described by the parton shower, avoiding all divergences. The two domains are claimed to be matched in a consistent and process-independent way.

The SHERPA [3] event generator offers an implementation of the CKKW algorithm. It has previously been shown to reproduce the shapes of NLO distributions [4] as predicted by the MC@NLO [5] and MCFM [6] programs to a high level of accuracy. A previous DØ study has compared SHERPA predictions with data for the variable $\Delta\phi$(jet,jet) in QCD di-jet events [7, 8]. Good agreement was observed over a range of four orders of magnitude.

In this note, a comparison is shown in the Z/γ^* + jet channel with $Z/\gamma^* \rightarrow e^+e^-$ between a Monte Carlo sample produced using SHERPA and data taken by the DØ experiment. A Monte Carlo sample produced using PYTHIA 6.314 [9], which generates jets using a parton shower approach, is included as a reference.

II. DATA AND MONTE CARLO SAMPLES

The data sample used in this analysis was collected between October 2002 and November 2005 by the DØ experiment at the Fermilab Tevatron collider at $\sqrt{s} = 1.96$ TeV. The integrated luminosity corresponds to about 950 pb^{-1}. Events containing two high p_T electrons in the final state were selected using a variety of single and di-electron triggers.

The SHERPA Monte Carlo sample was generated using the program version 1.0.6 and the CTEQ6L [10] parametrization of parton distribution functions (PDF). In the matrix element calculation up to three jets were included. The internal k_T algorithm parameter which separates jets described by the matrix element and those described by the parton shower was set to $(20 \text{ GeV})^2 / (1960 \text{ GeV})^2$. The PYTHIA sample was generated with PYTHIA 6.319 and the CTEQ6Ll parametrization of parton densities. For the underlying events model the parameter set corresponding to Tune A as described in Ref. [11] was used. In both Monte Carlo simulations zero bias events taken from data were overlayed on the generator events to account for additional activity from other $p\bar{p}$ interactions in the same beam crossing as the hard scatter.

Both Monte Carlo samples were processed through the full DØ detector simulation and reconstruction chain and were normalized to the total number of Z/γ^* events found in the data sample. No separate normalization was performed for exclusive subsamples.

III. EVENT SELECTION

The event selection starts by requiring two opposite charge electrons, both having a p_T above 25 GeV and being reconstructed within the pseudorapidity range $|\eta| < 2.5$. At least one of the electrons is required to be reconstructed in the central part of the detector ($|\eta| < 1.1$). To separate real electrons from the background from QCD jet production a likelihood function optimized for this purpose is used. The likelihood uses the following electron candidate properties as input: the fraction of the energy deposited in the electromagnetic layers of the calorimeter (EMF), the shower shape, the ratio E/p between the energy E measured in the calorimeter and the track momentum p, the spatial matching between the track and the calorimeter cluster, energy isolation and the distance of closest approach of the track to the primary vertex. The electrons are required to have an electromagnetic energy fraction larger than 90% and the di-electron invariant mass must be in the range 70 to 100 GeV.

Jet candidates are defined by the DØ Run II cone algorithm [12] and are required to have a transverse momentum $p_T > 15$ GeV. To suppress fake jets originating from instrumentation effects, a level-1 trigger confirmation is demanded and the candidate is required to have an EMF larger than 5%. Also, less than 46% of the detected jet energy should be deposited in the coarse hadronic layers of the calorimeter. To avoid electron contamination in the jet sample, the EMF of the electron candidate is required to be below 95% and all jet candidates must have a $\Delta\mathrm{R} = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ separation to the two electrons larger than 0.5.

An additional Gaussian smearing of the jet energies in Monte Carlo events was applied to account for differences between data and the Monte Carlo simulation. The width of the Gaussian smearing was determined from a study of the variable

$$\Delta S = \frac{(p_{T;projected}^{jet} - p_T^Z)}{p_T^Z} \tag{1}$$

in bins of p_t^Z for $Z + 1$ jet events, where $p_{T;projected}^{jet}$ is the projection of the transverse jet momentum to the axis defined by the transverse momentum of the Z candidate (reconstructed using the electron momenta).

IV. COMPARISON BETWEEN DATA AND MONTE CARLO

The aim of this study is to determine how accurately jet production in Z/γ^* events is modelled in the Monte Carlo generators SHERPA and PYTHIA. The jet description can either be probed directly, through a reconstruction of the jets, or indirectly by studying the p_T of the di-electron system which has to balance the p_T of the jet system.

In Fig. 1 the observed p_T distribution of the di-electron system is shown and compared to both the PYTHIA (left) and SHERPA (right) prediction. The shaded ranges in the histograms show the central value $\pm 1\sigma$ from Monte Carlo statistics. In the lower part of each plot the ratio between the number of events in data and the one predicted by the Monte Carlo models is given. An upward slope, corresponding to too few Z bosons with large p_T is found for PYTHIA, indicating a lack of hard jets. For SHERPA the agreement in the low p_T range is acceptable, however, at very large p_T the predicted spectrum appears to be too hard.

In Fig. 2 the observed jet multiplicity distribution is compared to the two predictions. The red band, labeled as *Pythia range stat*, indicates $\pm 1\sigma_{stat}$ for the Monte Carlo prediction. The red and blue bands combined, labeled as *Pythia range stat & sys*, show central values ± statistical and systematic errors combined in quadrature. Similarly, the grey data points are given with statistical errors only, whereas the black points represent the data with the combined statistical and the dominant systematic error, which results from uncertainties in the jet energy scale . The error on the ratio is found by Gaussian error propagation, treating the jet energy scale error of data and Monte Carlo as uncorrelated. Given that data and Monte Carlo share common sources of uncertainty, this is a conservative estimate.

The event numbers observed in data and predicted by the two Monte Carlo models are given in Table I. As mentioned above, the Monte Carlo predictions are normalized to the total number of events observed in data, however, no separate normalization is performed for the various jet Z + jet classes.

TABLE I: Number of events in the data for the different jet multiplicities in comparison to the predictioins from the Monte Carlos after normalization.

Sample	Inclusive	0-jet	1-jet	2-jet	3-jet	4-jet
Data	50417	40624	7877	1552	306	52
SHERPA	50417	39746	8410	1842	335	58
PYTHIA	50417	41271	7604	1324	193	23

For SHERPA, the central values of the predictions are somewhat higher than in data, whereas PYTHIA tends to produce too few multi-jet events. However, as seen in Fig. 2, within the large systematic uncertainties arising from low p_T jets, which dominate the distributions, the predictions of both Monte Carlo generators are in agreement with data.

The differential cross section $d\sigma/dp_T$ for the leading jet (Fig. 3), is consistent with the indications from the $p_T(\mathrm{Z})$ spectra. There is a clear, positive slope in the ratio for the PYTHIA prediction. The slope is found to be larger for the

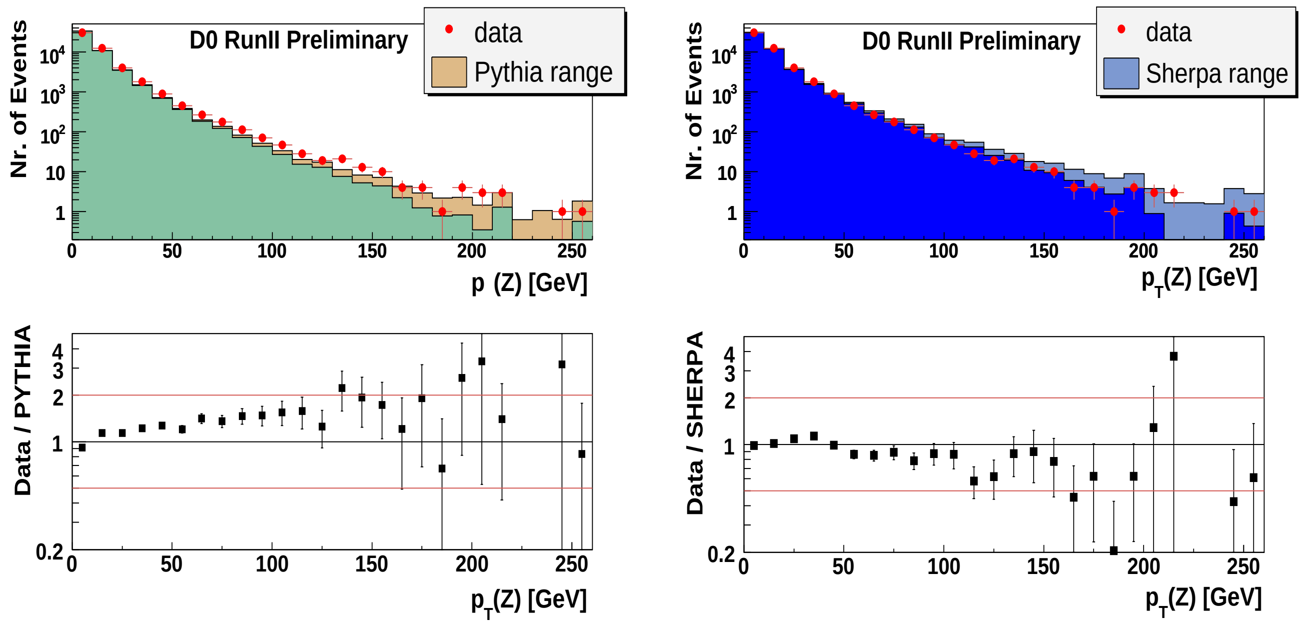

FIG. 1: The distribution of the transverse momentum of the Z boson, $p_T(\mathrm{Z})$: Data and PYTHIA (left), data and SHERPA (right). The lower plots show the ratio data / MC. The red lines indicate a factor 2 up and down.

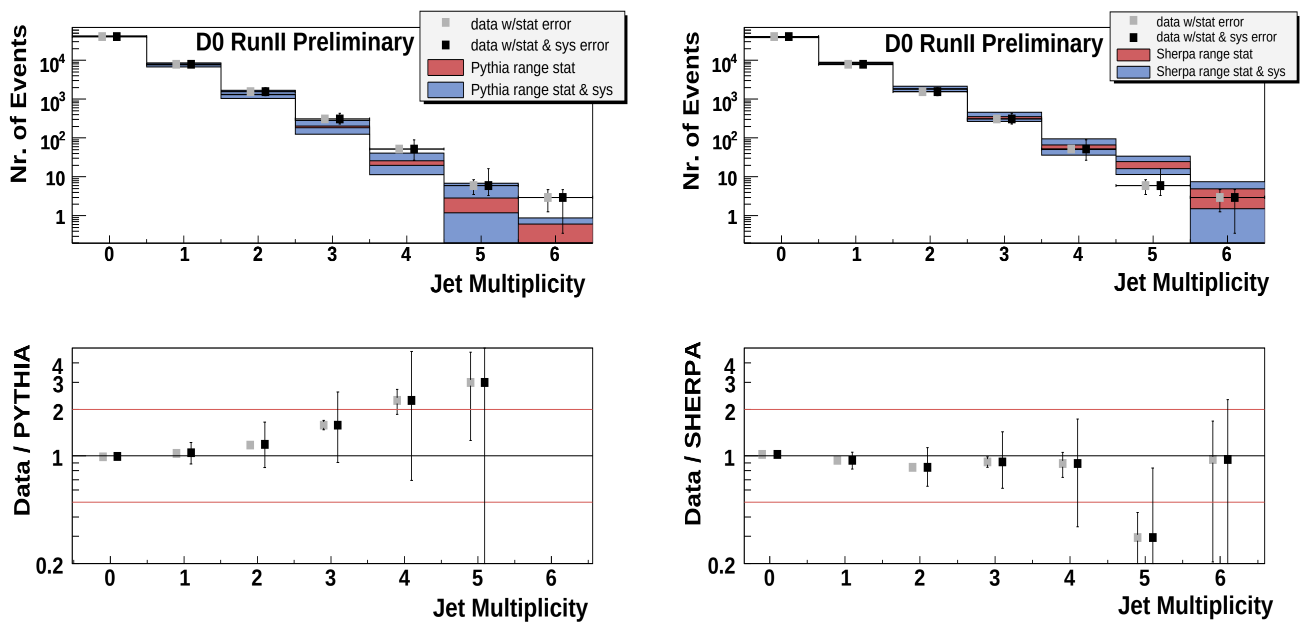

FIG. 2: Jet multiplicity: data and PYTHIA (left), data and SHERPA (right).

second and third jet (Figs. 4 and 5), leading, e.g., to a factor of more than 5 between data and the PYTHIA prediction at p_T=50 GeV for the third jet.

The SHERPA prediction for the p_T of the hardest jet, Fig. 3(right), is for most bins consistent with data within the systematic errors. The largest deviations, apart from the highest p_T bin where statistics are low, are found at around 80 GeV, where SHERPA predicts a factor 1.3 more jets than seen in data. The p_T spectra for the second and third hardest jets, Figs. 4 and 5, show an almost equally good match between SHERPA and data as seen for the hardest jet.

For most bins, SHERPA is consistent with data given the errors. The largest deviations for both the second and third hardest jets are seen at around 80 GeV where SHERPA predicts a factor 1.7 more jets than seen in data.

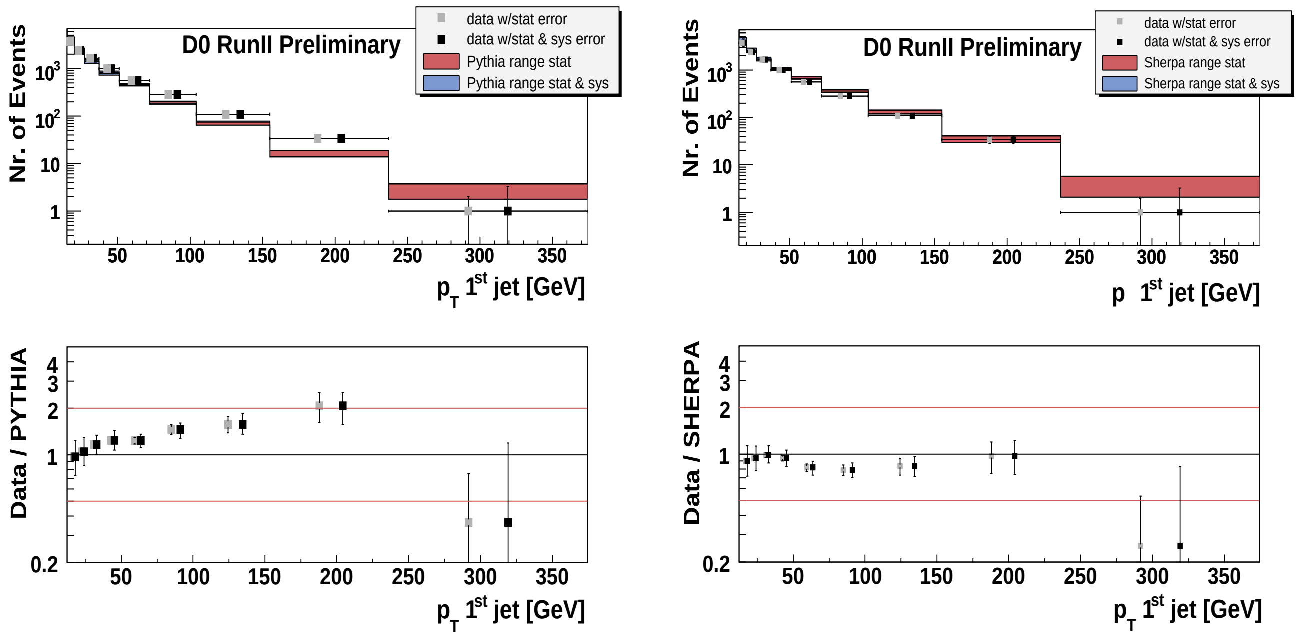

FIG. 3: p_T of hardest jet: data and PYTHIA (left), data and SHERPA (right).

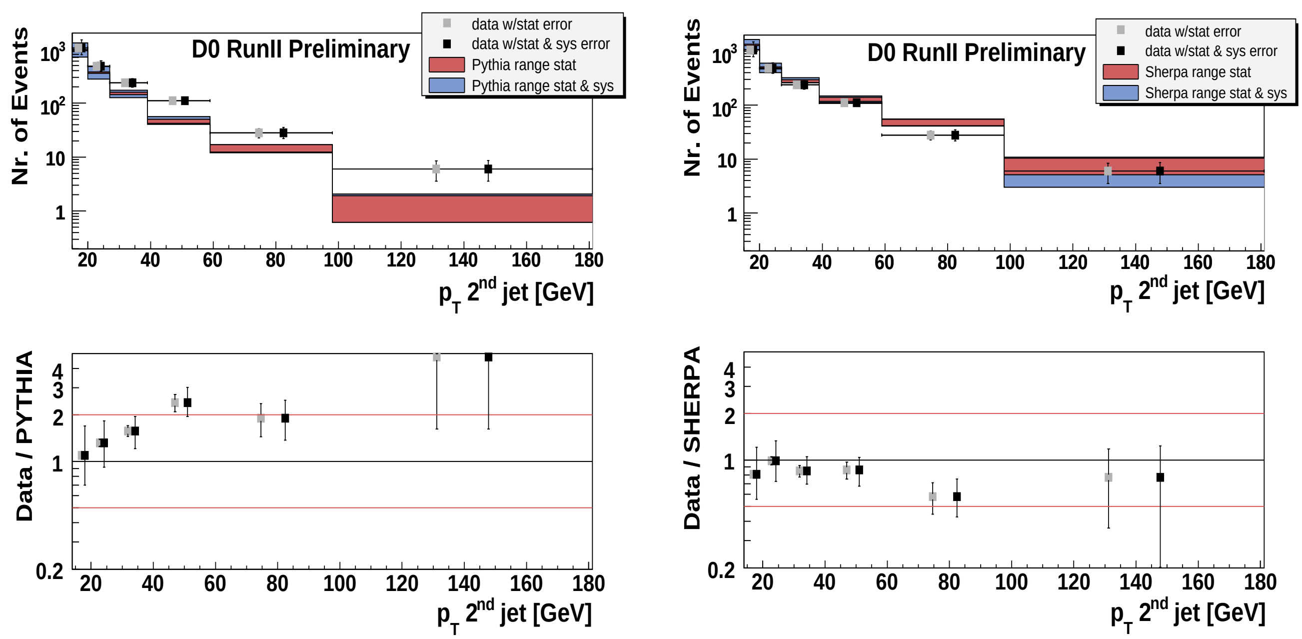

FIG. 4: p_T of second hardest jet: data and PYTHIA (left), data and SHERPA (right).

Angular correlations between pairs of hard final state jets with a large opening angle are expected to receive large interference contributions which are included in a matrix element description, but only partially modelled in parton shower approaches. It is therefore interesting to study the η and ϕ difference between the two hardest jets in events with at least two jets.

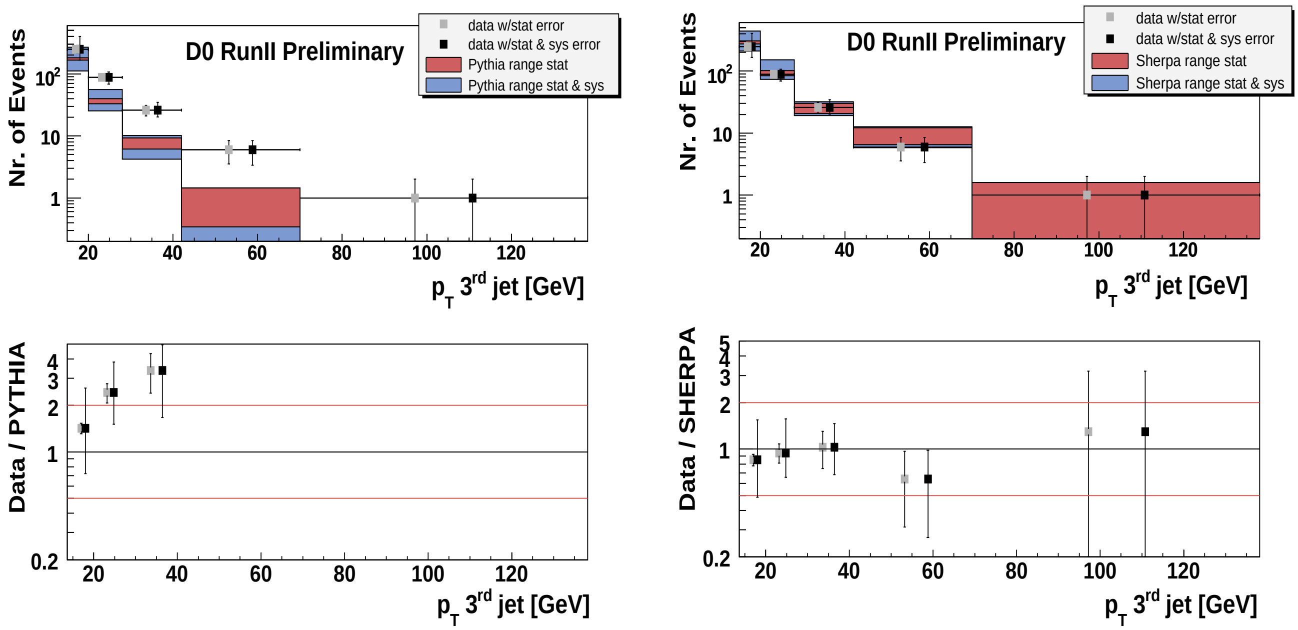

FIG. 5: p_T of third hardest jet: data and PYTHIA (left), data and SHERPA (right).

Both Monte Carlo generators offer a good description of the $\Delta\eta$(jet,jet) observable (Fig. 6). The distributions for $\Delta\phi(\mathrm{jet},\mathrm{jet})$ are shown in Fig. 7. SHERPA gives a very good description of data, and within the errors the ratio plot agrees with a straight line at 1. (It was seen in Fig. 2 that the SHERPA sample has about 20% more 2-jet events than data, but that this discrepancy is within the errors.) This agrees well with results from a DØ study of the same variable in di-jet events from QCD jet production [8]. Also the PYTHIA prediction for the shape of $\Delta\phi(\mathrm{jet},\mathrm{jet})$ distribution agrees well with data except at $\Delta\phi = \pi$ where a significant peak is seen for the PYTHIA prediction, which is, however, not observed in data. As for SHERPA, the overall normalization agrees with data within errors.

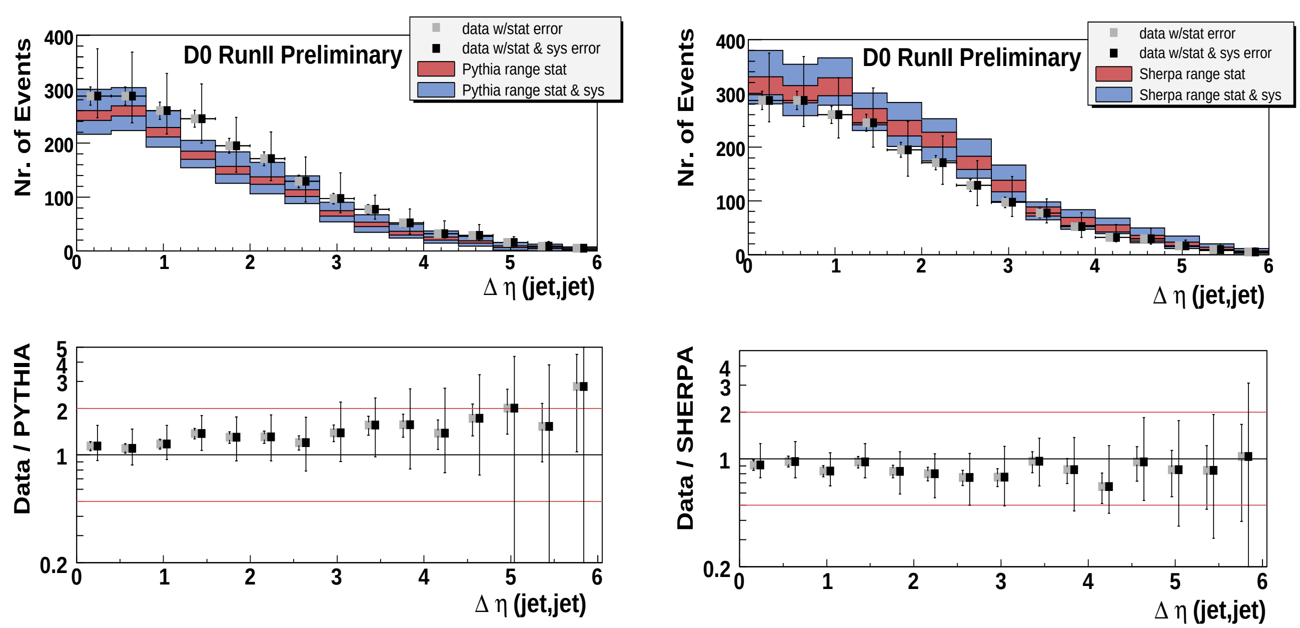

FIG. 6: $\Delta\eta(\mathrm{jet},\mathrm{jet})$: data and PYTHIA (left), data and SHERPA (right).

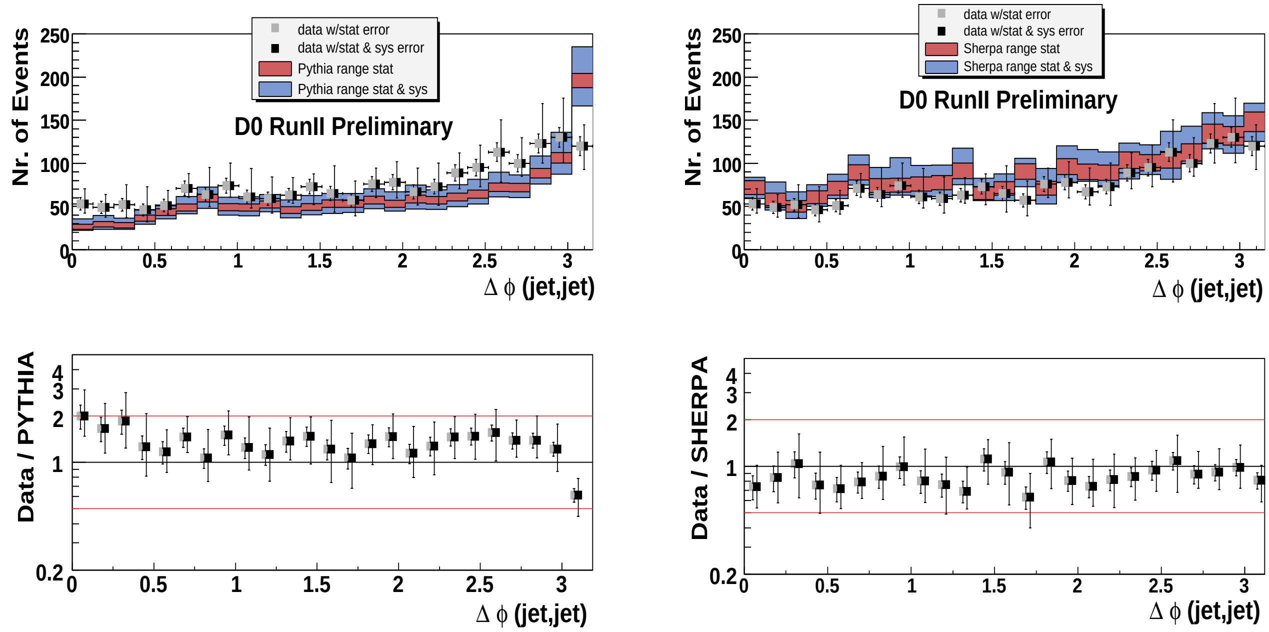

FIG. 7: $\Delta\phi(\text{jet}, \text{jet})$: data and PYTHIA (left), data and SHERPA (right).

An interesting class of events are those with three jets satisfying

$$|\eta_1 - \eta_2| > 2.0 \tag{2}$$

and

$$\eta_1 < \eta_3 < \eta_2 \quad \text{or} \quad \eta_2 < \eta_3 < \eta_1 \tag{3}$$

where η_i is η of the i^{th} hardest jet in the event. The purpose of studying these events is the similarity to the Vector Boson Fusion production channel of the Higgs boson [13]. This channel is characterized by one outgoing quark in each forward direction and a Higgs decaying in the central part of the detector. If the emission of QCD jets in the central part of the detector is correctly modelled by event generators, a veto on jet activity in the central part of the detector will be a powerful tool to suppress the large $t\bar{t}$ background.

The p_T distributions of the third hardest jet in events satisfying (2) and (3) are shown in Fig. 8. The description of η of the third jet, relative to the two hardest jets, is given by

$$\eta^* = \eta_3 - \frac{\eta_1 + \eta_2}{2}, \tag{4}$$

which is shown in Fig. 9. The number of events passing the tagging criteria is rather low, but SHERPA seems to describe both the overall rate and the shape of the distribution seen in data. PYTHIA gives a factor 1.7 fewer 3-jet events than data, and it therefore also gives too few events passing the tagging criteria. However, the ratio between data and the PYTHIA prediction is compatible with a straight line, so within errors the shape of the distribution is correctly described.

V. SYSTEMATIC UNCERTAINTIES

The main contribution to systematic uncertainties in this study comes from the uncertainty in the jet energy scale. Although the uncertainty on the jet energy is below 5% for most energies, the resulting uncertainty in the steeply falling p_T spectra is up to 25%.

Another main source of systematic uncertainty is the function used to smear the jet energies in Monte Carlo. Varying the jet smearing parameters by one standard deviation the largest deviation is observed in the low p_T bins of the jet p_T spectra, but even here the effect was smaller then 10%.

The effects of the systematic uncertainties arising from the jet energy scale and the smearing of jet energies in Monte Carlo have been propagated through to all distributions and were then added in quadrature.

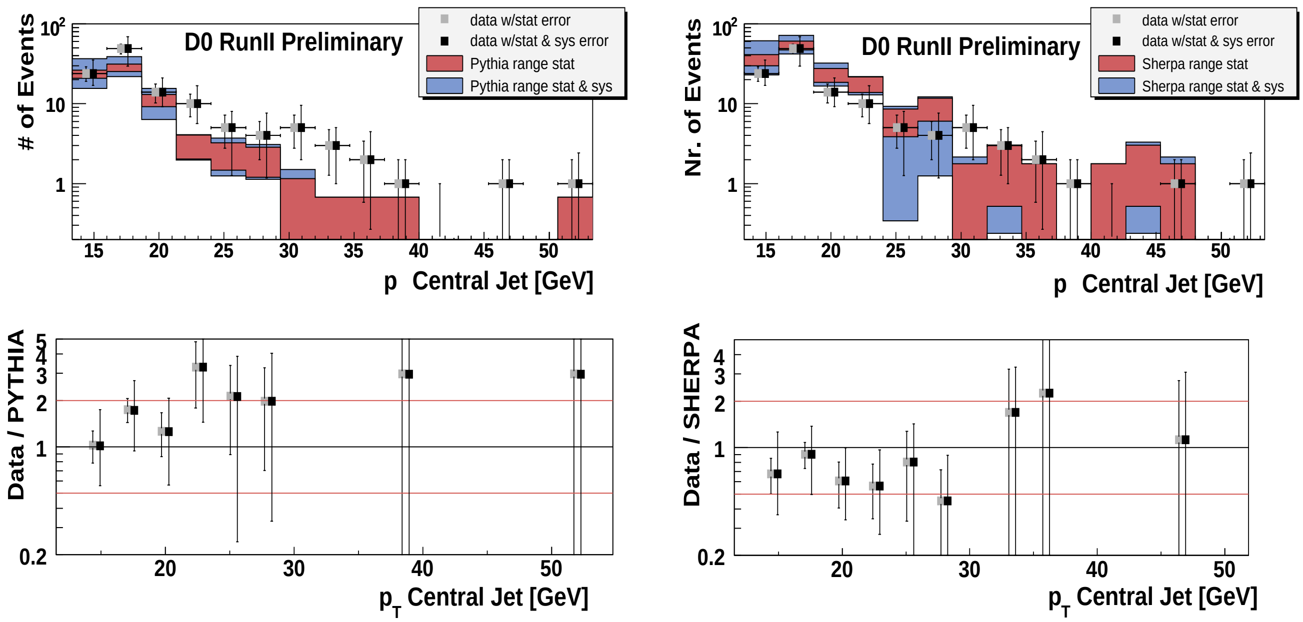

FIG. 8: p_T of third jet lying in between the two hardest jets in η.

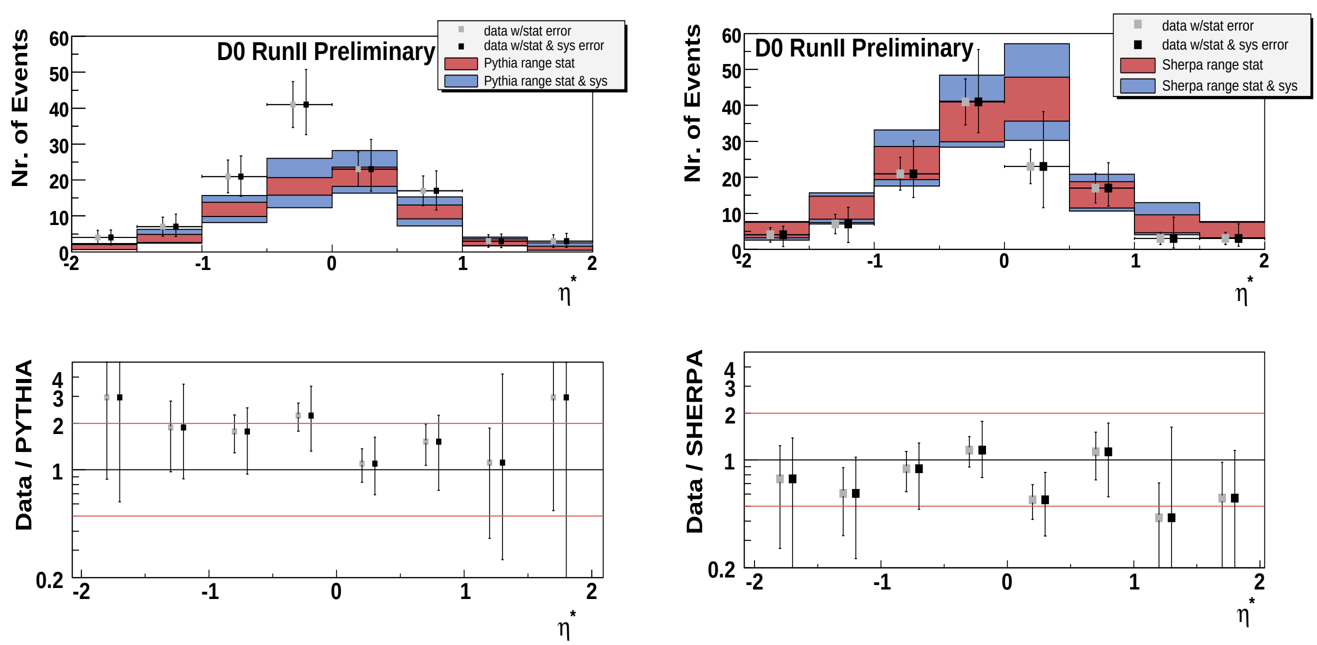

FIG. 9: η of third jet in the coordinate system where $\eta = 0$ lies in the middle of the two hardest jet.

VI. CONCLUSIONS

A first comparison between DØ data and the event generator SHERPA has been performed in the $Z/\gamma^* \rightarrow e^+e^-$ (+jets) channel, including a PYTHIA sample as reference.

The PYTHIA simulation describes jets through a parton shower algorithm which has been tuned to match a matrix element prediction for Z/γ^* + 1-jet production. The p_T spectra of the di-electron system and of the jets show that PYTHIA predicts fewer hard jets than seen in data, and the discrepancy increases with jet multiplicity. In the 2-jet sample, $\Delta\eta$ between the two hardest jets is well described. In the $\Delta\phi$(jet,jet) distribution a significant peak at π is seen in the PYTHIA sample but not in data.

The SHERPA Monte Carlo, combining parton shower and matrix element description of jets using the CKKW algorithm, has been found to offer a good description of jet properties. SHERPA has been shown to give an accurate description of jet multiplicities up four jets. Up to three jets were included in the matrix element when generating the sample. The p_T spectra predicted by SHERPA for the di-electron system as well as for the leading, second and third highest p_T jets are in reasonable agreement with the spectra observed in data. Also angular correlations between the two hardest jets in events with two or more jets are well described by SHERPA.

[1] S. Catani, F. Krauss, R. Kuhn and B. R. Webber, JHEP **0111** (2001) 063 [arXiv:hep-ph/0109231].
[2] F. Krauss, JHEP **0208**, (2002) 015 [arXiv:hep-ph/0205283].
[3] T. Gleisberg, S. Hoche, F. Krauss, A. Schalicke, S. Schumann and J. C. Winter, JHEP **0402** (2004) 056 [arXiv:hep-ph/0311263].
[4] F. Krauss, A. Schalicke, S. Schumann and G. Soff, Phys. Rev. D **70**, 114009 (2004) [arXiv:hep-ph/0409106].
[5] S. Frixione and B. R. Webber, JHEP **0206** (2002) 029 [arXiv:hep-ph/0204244].
[6] J. Campbell and R. K. Ellis, Phys. Rev. D **65** (2002) 113007 [arXiv:hep-ph/0202176].
[7] V. M. Abazov *et al.* [D0 Collaboration], Phys. Rev. Lett. **94** (2005) 221801 [arXiv:hep-ex/0409040].
[8] M. Zielinski, [arXiv:hep-exp/0602019].
[9] T. Sjöstrand, L. Lönnblad, S. Mrenna and P. Skands, [arXiv:hep-ph/0308153].
[10] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. Nadolsky and W. K. Tung, JHEP **0207** (2002) 012 [arXiv:hep-ph/0201195].
[11] R. Field, `http://www.phys.ufl.edu/~rfield/cdf/tunes/py_tuneA.html`
[12] G. C. Blazey *et al.*, arXiv:hep-ex/0005012.
[13] D. Zeppenfeld, "Connecting the Tevatron to Weak Boson Fusion", TeV4LHC plenary talk, `http://conferences.fnal.gov/tev4lhc/talks/040916_TeV4LHC_Plenary/99_Dieter_Zeppenfeld.pdf`.
[14] SHERPA homepage, `http://www.sherpa-mc.de/`.